7-2010

Nonparametric Testing for Asymmetric Information

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Martin SPINDLER
Ludwig Maximilians Universitat, Munchen

Nonparametric Testing for Asymmetric Information*

Liangjun Su[a], Martin Spindler[b]
[a] School of Economics, Singapore Management University, Singapore
[b] Munich Graduate School of Economics, Ludwig Maximilians University, Munich

July 28, 2010

Abstract

Asymmetric information is an important phenomenon in many markets and in particular in insurance markets. Testing for asymmetric information has become a very important issue in the literature in the last two decades. Almost all testing procedures that are used in empirical studies are parametric, which may yield misleading conclusions in the case of misspecification of either *functional* or *distributional* relationships among the variables of interest. Motivated by the literature on testing conditional independence, we propose a new nonparametric test for asymmetric information which is applicable in a variety of situations. We demonstrate the test works reasonably well through Monte Carlo simulations and apply it to an automobile insurance data set. Our empirical results consolidate Chiappori and Salanié's (2000) findings that there is no evidence for the presence of asymmetric information in the French automobile insurance market.

Keywords: Asymmetric information, Automobile insurance, Conditional independence, Distributional misspecification, Functional misspecification, Nonlinearity, Nonparametric test.

JEL classification codes: C12, C14, D82, D86, G22.

*Address Correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; E-mail: ljsu@smu.edu.sg, Phone: +65 6828 0386. Martin Spindler, Munich Graduate School of Economics, Kaulbachstrasse 45, 80539 Munich, Germany; E-mail: Martin.Spindler@lrz.uni-muenchen.de, Phone: +49 89 2180 1382. Spindler gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through GRK 801. We are very grateful to Bernard Salanié for his most valuable comments and for providing us with the data set.

1 Introduction

Since Akerlof (1970) the notion of asymmetric information, comprising adverse selection and moral hazard, has been explored at a rapid pace. At the same time people observed a wide gap between the theoretical development and empirical studies in asymmetric information. This gap has recently become narrower. In particular, the insurance market has been a fruitful and productive field for empirical studies. There are two reasons for this. First, insurance contracts are usually highly standardized and can exhaustively be described by a relatively small set of variables, and insurees' performances, i.e., the occurrence of a claim and possibly its cost, are exactly filed in the database of a insurance company. Second, insurance companies have hundreds of thousands or even millions of clients and therefore the samples are sufficiently large for econometric studies. Hence, fields like automobile insurance, annuities and life insurance, crops insurance, long-term care and health insurance offer a large sample of standardized contracts for which performances are recorded and therefore are well suited for testing the theoretical predictions of contract theory. For a detailed justification for using insurance data to test contract theory, see Chiappori and Salanié (1997). For a recent overview over the issue of testing for adverse selection in insurance markets, see Cohen and Siegelman (2010). The latter paper covers a large number of empirical studies in different insurance branches.

In statistical terms, the theoretical notion of asymmetric information implies a positive (conditional) correlation between coverage and risk. In their seminal paper Chiappori and Salanié (2000) propose both parametric and nonparametric methods to test this. Their non-parametric tests are restricted to discrete data with only two categories per variable even though some of the variables in the data set are continuous and others have far more than two categories. Therefore, in order to conduct Chiappori and Salanié's nonparametric test, all variables must be transformed to binary variables, which often results in a loss of information. The implication of such transformation has not been clear to us yet. Following the lead of Chiappori and Salanié (2000), most subsequent studies use a variation of their parametric testing procedure which has become somewhat standard in the empirical contract theory. Nevertheless, these parametric tests are fragile to both *functional* and *distributional* form misspecifications which are a severe problem in this field. For example, in the automobile insurance market it is common knowledge that the age of the driver has a nonlinear effect on the probability of an accident, but such a nonlinear effect has rarely been taken into account in the literature. For another example, the error term in the binary model for modeling the choice of an insurance contract may not be either normally or logistically distributed, and tests for asymmetric information based on the probit or logit model can therefore yield misleading conclusions in the case of incorrect distributional specification. For this reason, in this paper we propose a new purely nonparametric test for asymmetric information based on the notion of conditional independence, which avoids the problem of either functional or distributional misspecification.

The absence of asymmetric information means that the choice of a contract Y (discrete variable) provides no information for predicting the "performance" variable Z (discrete or continuous, e.g., the number of claims or the sum of reimbursements), conditional on the vector X of all exogenous variables (discrete and continuous). Therefore we can transform the problem of testing the absence of asymmetric information into a test for conditional independence: $F(Z|X,Y) = F(Z|X)$ almost surely (a.s.) where, e.g., $F(Z|X,Y)$ denotes the conditional cumulative distribution function (CDF) of Z given (X,Y). We propose a nonparametric test statistic to test the conditional independence of Z and Y given X. We show that the test statistic is asymptotic normally distributed under the null hypothesis of conditional independence (or absence of asymmetric information) and diverges to infinity in the presence of conditional dependence (or asymmetric information). We then apply our test to a French automobile insurance data set and compare our testing results with the results found in the literature.

The rest of the paper is structured as follows. Section 2 outlines the theory of asymmetric information. Section 3 reviews the standard statistical tools for testing asymmetric information. We introduce a new nonparametric test for conditional independence in Section 4. We conduct a small set of Monte Carlo simulations to examine the performance of the new test in Section 5. We apply our test to test for the asymmetric information in the French insurance market in Section 6. Final remarks are contained in Section 7. All technical details are relegated to the Appendix.

2 The Theory of Asymmetric Information

In their seminal paper Rothschild and Stiglitz (1976) introduce the notion of adverse selection in insurance markets that has since then been extended in many directions.[1] In the basic model, the insureds have private information about the expected claim, exactly speaking about the probability that a claim with fixed level occurs, while the insurers do not have this information. Thus there are two groups with different claim probabilities, the "bad" and

[1] For a detailed survey on adverse selection and the related moral hazard problem, see Dionne, Doherty and Fombaron (2000) and Winter (2000), respectively.

"good" risks. The agents have identical preferences which are moreover perfectly known to the insurer. Additionally, perfect competition and exclusive contracts are assumed. Exclusive contracts mean that an insuree can buy coverage only from one insurance company. This allows firms to implement nonlinear (especially convex) pricing schemes which are typical under asymmetric information. Under this setting insurance companies offer a menu of contracts in equilibrium: a full insurance which is chosen by the "bad" risks and a partial coverage which is bought by the "good" risks. In general, contracts with more comprehensive coverage are sold at a higher (unitary) premium.

Clearly, one expects a positive correlation between "risk" and "coverage" (conditional on observables). Since the assumptions in the Rothschild and Stiglitz model are very simplistic and normally not fulfilled in real applications, an important question to address is how robust this coverage-risk correlation is. Chiappori et al. (2006) show that the positive correlation property extends to much more general models, as already conjectured by Chiappori and Salanié (2000). Under competitive markets this property is also valid in a very general framework entailing heterogeneous preferences, multiple level of losses, multidimensional adverse selection plus possible moral hazard and even non-expected utility theory. In the case of imperfect competition some form of positive correlation must hold if the agent's risk aversion becomes public information. In the case of private information the property does not necessarily hold (c.f. Jullien et al. (2007)).

While adverse selection concerns "hidden information", moral hazard deals with "hidden action". Moral hazard occurs when the expected loss (accident probability or level of damage) is not exogenous, as assumed in the adverse selection case, but depends on some decision or action made by the subscriber (e.g., effort or prevention) which is neither observable nor contractible. A higher coverage leads to decreased efforts and therefore to a higher expected loss. Therefore moral hazard also predicts a positive correlation between "coverage" and "risk".

Although both phenomena lead to a positive risk-coverage correlation, there is one important difference: under adverse selection the risk of the potential insuree affects the choice of the contract, whereas under moral hazard the chosen contract influences the behavior and therefore the expected loss. So there exists reversed causality in both cases.[2]

In sum, the theory of asymmetric information[3] predicts a positive correlation between

[2] To disentangle moral hazard from adverse selection is an important problem in the empirical literature. The first attempt is Dionne et. al. (2004). An overview over different possible strategies for dealing with this problem can be found in Cohen and Siegelman (2010).

[3] It seems that in the empirical insurance literature adverse selection is more stressed than the moral hazard aspect which only receives minor attention, see, e.g., Cohen and Siegelman (2010).

(appropriately defined) "risk" and "coverage" which should be quite robust.

To proceed, it is worth mentioning that to test for asymmetric information, the researcher needs to access to the same information which is also available to the insurer and used for pricing. The theory of adverse selection predicts that the insurance company offers a menu of contracts to indistinguishable individuals. Individuals are (ex ante) indistinguishable for the insurer if they share the same characteristics. Therefore the positive risk-coverage correlation is valid only conditional on the observed characteristics. Different groups of observable equivalent individuals are offered different menus of contracts with different prices according to their risk exposure.[4] Only within each class are the mechanisms described above valid.

3 Standard Testing Procedures

In this section we review some tests of asymmetric information in the literature. We first outline the general structure of the problem, and then review the parametric and nonparametric testing procedures in turn.

3.1 General Structure

In the following we denote by X the vector of exogenous control variables to be conditional on, by Y a decision or choice variable, and by Z the endogenous "performance" variable. In the context of insurance, X usually includes variables that are used for risk classification by the insurance company, Y could be the choice of deductibles, and Z could be the number of accidents or claims[5] or the sum of reimbursements caused by accidents. As we shall see, we allow both continuous and discrete variables in X, and Z can be continuous or discrete. For concreteness, we assume that Y is a discrete variable. There is no asymmetric information if and only if the prediction of the endogenous variable Z based on X and Y jointly coincides with its prediction based on X alone. Formally, this can be stated in terms of the equivalence of two conditional CDFs:

$$F(Z|X,Y) = F(Z|X) \text{ a.s.,} \tag{3.1}$$

where, e.g., $F(Z|X,Y)$ denotes the conditional CDF of Z given (X,Y). Intuitively, this means that the the choice of the contract, e.g., the choice of a certain deductible, provides no useful information for predicting the risk, e.g., the number of claims, as soon as the risk classes are

[4] For the theory of risk classification under asymmetric information see Crocker and Snow (2000).

[5] The distinction of accidents and claims is a very important point in the empirical literature as not every accident leads to a claim. Neglecting this issue might lead to biased results.

controlled for. Equivalently, we can interchange the roles of Z and Y :

$$F(Y|X,Z) = F(Y|X) \text{ a.s.,} \tag{3.2}$$

where, e.g., $F(Y|X,Z)$ denotes the conditional CDF of Y given (X,Z). (3.2) says that the number of claims (or the sum of reimbursements caused by accidents) does not provide useful information to predict the choice of deductibles as long as we control the risk classes. Either (3.1) or (3.2) indicates the conditional independence of Y and Z given X.[6]

3.2 Parametric Testing Procedures

Almost all empirical studies analyzing the positive risk-coverage correlation property use one of the following two types of parametric procedures.

The first approach is to run a regression of Z_i on Y_i and X_i and test whether the coefficient of Y_i is zero or not. When Z_i is continuously valued, the regression model is

$$Z_i = \beta_0 + \beta_1 Y_i + \beta_2' X_i + \varepsilon_i, \tag{3.3}$$

where ε_i is the error term, β_0, and (β_1, β_2') are intercept and slope coefficients, respectively, and the prime denotes transpose. When Z_i is a dummy variable, the regression model is

$$Z_i = \mathbf{1}(\beta_0 + \beta_1 Y_i + \beta_2' X_i + \varepsilon_i > 0) \tag{3.4}$$

where ε_i is assumed to be either normally or logistically distributed, and $\mathbf{1}(A) = 1$ if A is true and 0 otherwise. If Z_i is a discrete variable that has more than two categories, then one can use the ordered logit model. One obvious drawback of this approach is that it neglects by construction the potential nonlinear effects of the controlled variables, and a test based on (3.3) is designed to test the conditional mean independence of Z_i and Y_i given X_i, which is a much weaker condition than conditional independence at the distributional level. In addition, the distributional assumption in the probit, logit, or ordered logit model may not hold, and once this happens, tests for asymmetric information can lead to misleading conclusions.

In one of the first empirical studies Puelz and Snow (1994) consider an ordered logit formulation for the deductible choice variable and find strong evidence for the presence of asymmetric information in the market for automobile collision insurance in Georgia. But

[6] Alternatively, one can use conditional probability density or mass functions to form the independence between Y and Z conditional on X : $f(Z|X,Y) = f(Z|X)$, or $f(Y|X,Z) = f(Y|X)$ a.s., where e.g., $f(Z|X,Y)$ denotes the conditional probability density or mass function of Z given (X,Y). See Su and White (2007, 2008, 2010) for other equivalent formulations.

Dionne et al. (2001) show that this correlation might be spurious because of the highly constrained form of the exogenous effects or the misspecification of the functional form used in the regression. They propose to add the estimate $\hat{E}(Z_i|X_i)$ of the conditional expected value of Z_i given X_i as a regressor into the ordered logit model to take into account the nonlinear effect of the risk classification variables, and by accounting for this, they find no residual asymmetric information in the market for Canadian automobile insurance.

A second and more advanced approach was introduced by Chiappori and Salanié (1997, 2000) and has become widespread in the empirical contract theory since then. They define two probit models, one for the choice of the coverage Y_i (either compulsory/basic coverage or comprehensive coverage) and the other for the occurrence of an accident Z_i (either no accident being blamed for or at least one accident with fault):

$$\begin{cases} Y_i = \mathbf{1}(\beta' X_i + \varepsilon_i > 0) \\ Z_i = \mathbf{1}(\gamma' X_i + \eta_i > 0) \end{cases} \tag{3.5}$$

where ε_i and η_i are independent standard normal errors, and β and γ are coefficients. They first estimate these two probit models independently, calculate the generalized residuals $\hat{\varepsilon}_i$ and $\hat{\eta}_i$, [7] and then construct the following test statistic

$$W_n = \frac{\left(\sum_{i=1}^n \hat{\varepsilon}_i \hat{\eta}_i\right)^2}{\sum_{i=1}^n \hat{\varepsilon}_i^2 \hat{\eta}_i^2}. \tag{3.6}$$

Under the null of conditional independence, $\text{cov}(\varepsilon_i, \eta_i) = 0$ and W_n is distributed asymptotically as $\chi^2(1)$. Alternatively, one can estimate a bivariate probit model in which ε_i and η_i are distributed as bivariate normal with correlation coefficient ρ to be estimated, and then test whether $\rho = 0$ or not. They find no evidence of asymmetric information in the French automobile insurance market.

3.3 Nonparametric Testing Procedures

Motivated by the χ^2-test for independence in the statistics literature, Chiappori and Salanié (2000) propose a nonparametric test for asymmetric information by restricting all variables in X_i, Y_i, and Z_i to be binary. They choose a set of m exogenous binary variables in X_i, and construct $M \equiv 2^m$ cells in which all individuals have the same values for all variables in X_i. For each cell they set up a 2×2 contingency table generated by the binary values of Y_i and

[7] For example, the generalized residual $\hat{\varepsilon}_i$ estimates $E(\varepsilon_i|Y_i)$. See Gourieroux et al. (1987) for the definition of generalized residuals in limited dependent models.

Z_i, and conduct a χ^2-test for independence. This results in M test statistics, each of which is distributed asymptotically as $\chi^2(1)$ under the null hypothesis. They aggregate these M test statistics in three ways to obtain three overall test statistics for conditional independence: one is the Kolmogorov-Smirnoff test statistic that compares the empirical distribution function of the M test statistics with the CDF of the $\chi^2(1)$ distribution; the second is to count the number of rejections for the independence test for each cell which is asymptotically distributed as binomial $B(M, \alpha)$ under the null, where α denotes the significance level of the χ^2 test within each cell; and the third is the sum of all the test statistics for each individual cell, which is asymptotically $\chi^2(M)$ distributed under the null. Again, using these nonparametric methods, they find no evidence for the presence of asymmetric information in the French automobile insurance market.

4 A New Nonparametric Test

In this section we propose a new nonparametric test for asymmetric information based on the formulation in (3.1). The null hypothesis is

$$H_0 : F(Z|X, Y) = F(Z|X) \text{ a.s.,} \tag{4.1}$$

and the alternative hypothesis is

$$H_1 : \Pr\{F(Z|X, Y) = F(Z|X)\} < 1. \tag{4.2}$$

We consider the case where Y is a discrete random variable (typically dummy), Z can be either discrete or continuous, and X contains both continuous and discrete variables. Note that early literature on testing for conditional independence mainly focus on the case where both Y and X are continuously distributed, see, Delgado and González-Manteiga (2001), Su and White (2007, 2008, 2010), Song (2009), Huang (2009), Huang and White (2009), to name just a few. Even though we restrict our attention mainly on the case where Y is discrete, we remark that in the case of continuous Y, the proposed test continues to work with little modification.

4.1 The Test Statistic

Given observations $\{(X_i, Y_i, Z_i)\}_{i=1}^n$, one could propose a test based on the comparison of two conditional cumulative distribution (CDF) estimates, one is the conditional CDF of Z given X

$(F(z|x))$ and the other is the conditional CDF of Z given (X,Y) $(F(z|x,y))$.[8] Nevertheless, for the reason elaborated at the end of this section, we will compare $F(z|x,y)$ with $F(z|x,\widetilde{y})$ for different values y and $\widetilde{y}$ instead.

To allow for both continuous and discrete regressors in X_i, write $X_i = (X_i^{c\prime}, X_i^{d\prime})'$ where X_i^c denotes a $p_c \times 1$ vector of continuous regressors in X_i and X_i^d denotes a $p_d \times 1$ vector of remaining discrete regressors with $p_d \equiv p - p_c$. For simplicity, we assume that none of the discrete regressors has a natural ordering and each takes only a finite number of values.[9] We use X_{is}^c (X_{is}^d) to denote the sth component of X_i^c (X_i^d), where $s = 1, \cdots, p_c$ (p_d). We assume that X_{is}^d takes c_s different values in $\mathcal{X}_s^d \equiv \{0, 1, \cdots, c_s - 1\}$, $s = 1, \cdots, p_d$, and Y_i takes c_y different values in $\mathcal{Y} \equiv \{0, 1, \cdots, c_y - 1\}$.

Fix $y \in \mathcal{Y}$. We consider the estimation of $F(z|x,y)$ by using the local linear method. For this purpose, we define the kernels for the continuous regressor X_i^c and discrete regressor X_i^d separately. For the continuous regressor, we choose a product kernel function $Q(\cdot)$ of $q(\cdot)$ and a vector of smoothing parameters $h \equiv (h_1, ..., h_{p_c})'$. Let $Q_{h,j}(x^c) \equiv \Pi_{s=1}^{p_c} h_s^{-1} q\left(\left(X_{js}^c - x_s^c\right)/h_s\right)$ and

$$Q_{h,ji} \equiv Q_h\left(X_j^c - X_i^c\right) = \prod_{s=1}^{p_c} h_s^{-1} q\left(\left(X_{js}^c - X_{is}^c\right)/h_s\right), \tag{4.3}$$

where, for example, $x^c \equiv (x_1^c, \cdots, x_{p_c}^c)'$, and X_{is}^c denote the sth element in X_i^c. For the discrete regressor, we follow Racine and Li (2004) and Li and Racine (2007, 2008) and use a variation of the kernel function of Aitchison and Aitken (1976):

$$l\left(X_{js}^d, X_{is}^d, \lambda_s\right) = \begin{cases} 1 & \text{if } X_{js}^d = X_{is}^d \\ \lambda_s & \text{otherwise} \end{cases} \tag{4.4}$$

where $\lambda_s \in [0,1]$ is the smoothing parameter. In the special case where $\lambda_s = 0$, $l(\cdot,\cdot,\cdot)$ reduces to the usual indicator function as used in the nonparametric frequency approach. Similarly, $\lambda_s = 1$ leads to a uniform weight function, in which case, the X_{is}^d regressor will be completely smoothed out in the sense that it will not affect the nonparametric estimation result. The

[8] For more rigorous notation, one could use $F_{Z|X}(z|x)$ $(F_{Z|X,Y}(z|x,y))$ to denote the conditional CDF of Z given X ((X,Y)). Below we make reference to these CDFs and several probability density functions (PDFs) simply using the list of their arguments – for example, $p(x,y,z)$, $p(x,y)$ and $p(x)$ denote the PDFs of (X_i, Y_i, Z_i), (X_i, Y_i), and X_i, respectively. This notation is compact, and we hope, sufficiently unambiguous. In addition, even though a PDF is most commonly associated with continuous distributions, here we use it to denote the Radon–Nikodym derivative of a CDF with respect to the Lebesgue measure for the continuous component and the counting measure for the discrete component.

[9] When some of the conditioning variables in X_i have a natural ordering, one can easily modify the discrete kernel defined below following either Racine and Li (2004) or Li and Racine (2007, 2008).

product kernel function for all the discrete vector is given by

$$L_{\lambda,ji} \equiv L_\lambda\left(X_j^d, X_i^d\right) \equiv \prod_{s=1}^{p_d} \lambda_s^{\mathbf{1}(X_{js}^d \neq X_{is}^d)}, \tag{4.5}$$

where $\lambda \equiv (\lambda_1, \cdots, \lambda_{p_d})'$. Combining (4.3) and (4.5), we obtain the product kernel function for the conditioning vector X_i:

$$K_{h\lambda,ji} \equiv K_{h\lambda}(X_j, X_i) = Q_h\left(X_j^c - X_i^c\right) L_\lambda\left(X_j^d, X_i^d\right). \tag{4.6}$$

Now, fix a point $X_i = (X_i^{c\prime}, X_i^{d\prime})'$. It follows from the first order Taylor expansion that

$$F(z|X_j, y) \approx F(z|X_i, y) + \dot{F}(z|X_i, y)'\left(X_j^c - X_i^c\right) \tag{4.7}$$

for any X_j^c in the neighborhood of X_i^c and $X_j^d = X_i^d$, where $\dot{F}(z|x, y) = \partial F(z|x)/\partial x^c$, i.e., the derivative is only taken with respect to the continuous component x^c of $x \equiv (x^{c\prime}, x^{d\prime})'$. Given observations $\{(X_i, Y_i, Z_i)\}_{i=1}^n$, we estimate $F(Z_i|X_i, y)$ by solving the weighted least squares minimization problem

$$\min_{\boldsymbol{\beta}} \sum_{j=1}^n \left[\mathbf{1}\{Z_j \leq Z_i\} - \beta_0 - \beta_1'\left(\left(X_j^c - X_i^c\right)/h\right)\right]^2 K_{h\lambda,ji}\mathbf{1}_j^y, \tag{4.8}$$

where $\boldsymbol{\beta} \equiv \left(\beta_0, \beta_1'\right)'$ and $\mathbf{1}_j^y \equiv \mathbf{1}(Y_j = y)$. Our estimator $\widehat{F}(Z_i|X_i, y)$ is the minimizing intercept term in the above problem. Let $\boldsymbol{\tau}_h\left(X_j^c - x^c\right) \equiv \left(1, \left(\left(X_j^c - x^c\right)/h\right)'\right)'$. Then it is easy to verify that

$$\widehat{F}(Z_i|X_i, y) = e_1' \left[\mathbf{S}_{ny}(X_i)\right]^{-1} \frac{1}{n}\sum_{j=1}^n K_{h\lambda,ji}\mathbf{1}_j^y \boldsymbol{\tau}_h\left(X_j^c - X_i^c\right) \mathbf{1}(Z_j \leq Z_i)$$

where $e_1 \equiv (1, 0, \cdots, 0)'$ is a $(p_c + 1)$-vector, and $\mathbf{S}_{ny}(X_i) \equiv \frac{1}{n}\sum_{j=1}^n K_{h\lambda,ji}\mathbf{1}_j^y \boldsymbol{\tau}_h(X_j^c - X_i^c)\boldsymbol{\tau}_h(X_j^c - X_i^c)'$.

We measure the variations in $\widehat{F}(Z_i|X_i, y)$ across different values of y and different observations by

$$D_n \equiv \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^n \left[\widehat{F}(Z_i|X_i, r) - \widehat{F}(Z_i|X_i, s)\right]^2.$$

We study the asymptotic properties of D_n under H_0, a sequence of Pitman local alternatives, and the global alternative H_1. We will show that after being appropriately recentered and scaled, D_n is asymptotically normally distributed under the null and local alternatives, and diverges to infinity under the global alternative.

4.2 Assumptions

Throughout the paper we use ξ_i, ζ_i, and ς_i to denote $(X_i', Y_i, Z_i)'$, $(X_i', Y_i)'$, and $(X_i', Z_i)'$, respectively. Similarly, let $\xi \equiv (x', y, z)'$, $\zeta \equiv (x', y)'$ and $\varsigma \equiv (x', z)'$. With a little bit abuse of notation, we use $p(\xi)$, $p(\zeta)$, and $p(x)$ denote the PDF of ξ_i, ζ_i, and X_i, respectively. Similarly, $F(z|x,y) \equiv F(z|x^c, x^d, y)$ denotes the conditional CDF of Z_i given $(X_i^{c\prime}, X_i^{d\prime}, Y_i)'$.

To facilitate our asymptotic analysis, we make the following assumptions.

Assumption A.1 The sequence $\{\xi_i\}_{i=1}^n$ is independent and identically distributed (IID) with CDF F_ξ.

Assumption A.2 (i) The support $\mathcal{X}^c$ of X_i^c is compact.

(ii) $p(\xi)$ is uniformly bounded over its support $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, where $\mathcal{X} \equiv \mathcal{X}^c \times \mathcal{X}^d$, $\mathcal{X}^d \equiv \mathcal{X}_1^d \times \cdots \times \mathcal{X}_{p_d}^d$, and $\mathcal{Z}$ is the support of Z_i. $p(\zeta) \equiv p\left(x^c, x^d, y\right)$ is bounded away from 0 for all $x^c \in \mathcal{X}^c$, $x^d \in \mathcal{X}^d$, and $y \in \mathcal{Y}$.

Assumption A.3 Let $\eta \equiv \left(x^d, y\right)$. (i) For each $\eta \in \mathcal{X}^d \times \mathcal{Y}$ and $z \in \mathcal{Z}$, $F(z|x^c, \eta)$ is Lipschitz continuous in $x^c \in \mathcal{X}^c$ and has all partial derivatives up to order 2 with respect to x^c.

(ii) For each $\eta \in \mathcal{X}^d \times \mathcal{Y}$ and $z \in \mathcal{Z}$, the second order partial derivatives with respect to x^c, $\partial^2 F(z|x^c, \eta)/\partial x_s^c \partial x_t^c$, $s, t = 1, \cdots, p_c$, are uniformly bounded and Hölder continuous on $\mathcal{X}^c$: for x^c, $\tilde{x}^c \in \mathcal{X}^c$, $|\partial^2 F(z|x^c, \eta)/\partial x_s^c \partial x_t^c - \partial^2 F(z|\widetilde{x}^c, \eta)/\partial x_s^c \partial x_t^c| \leq C \; ||x^c - \widetilde{x}^c||$, where C is a generic finite constant and $\|\cdot\|$ denotes the Euclidean norm.

(iii) For each $x^c \in \mathcal{X}^c$ and $\eta \in \mathcal{X}^d \times \mathcal{Y}$, $|F(z|x^c, \eta) - F(\widetilde{z}|x^c, \eta)| \leq C \; |z - \widetilde{z}|$ for all $z, \widetilde{z} \in \mathcal{Z}$.

Assumption A.4 (i) The kernel function $q : \mathbb{R} \to \mathbb{R}^+$ is a continuous, bounded, and symmetric PDF.

(ii) $u \to |u|^4 q(u)$ is integrable on $\mathbb{R}$ with respect to Lebesgue measure.

(iii) Let $\mathbf{q}_j(u) \equiv u^j q(u)$ for all $j = 0, \cdots, 3$. For some $C_1 < \infty$ and $C_2 < \infty$, either $q(\cdot)$ is compactly supported such that $q(u) = 0$ for $|u| > C_1$, and $|\mathbf{q}_j(u) - \mathbf{q}_j(\widetilde{u})| \leq C_2 |u - \widetilde{u}|$ for any $u, \widetilde{u} \in \mathbb{R}$ and for all $j = 0, \cdots, 3$; or $q(\cdot)$ is differentiable, $|d\mathbf{q}_j(u)/du| \leq C_1$, and for some $\iota_0 > 1$, $|d\mathbf{q}_j(u)/du| \leq C_1 |u|^{-\iota_0}$ for all $|u| > C_2$ and for all $j = 0, \cdots, 3$.

Assumption A.5 Let $h! \equiv \Pi_{s=1}^{p_c} h_s$. As $n \to \infty$, $\|h\| \to 0$, $\|\lambda\| \to 0$, $||\lambda||$ *is of the same order as* $||h||^2$, $n(h!)^2/\log n \to \infty$, $n(h!)^{1/2}\|h\|^4 \to 0$, and $\|h\|^4/h! \to 0$.

Remark 1. The IID assumption in assumption A.1 is standard in cross sectional study. One could allow heterogeneity but that would complicate the presentation to a large degree. Assumption A.2 is standard for nonparametric local polynomial estimation with mixed regressors. Assumptions A.3-A.4 are used to obtain uniform consistency for the local polynomial estimator of Masry (1996) and Hansen (2008). Assumption A.5 imposes appropriate conditions on the bandwidth. In particular A.5 implies that undersmoothing is required for our test and $p_c < 4$. This is typical in nonparametric tests when local linear regression is involved. In the case where $p_c \geq 4$, one has to rely upon higher order local polynomial regressions.

4.3 The Asymptotic Distribution of the Test Statistic

Let $\overline{\mathbf{S}}_y(x) \equiv E[K_{h\lambda}(X_j, x)\,\mathbf{1}_j^y \boldsymbol{\tau}_h(X_j^c - x^c)\boldsymbol{\tau}_h(X_j^c - x^c)'], \mathbf{K}_y(\zeta_j, x) \equiv e_1'[\overline{\mathbf{S}}_y(x)]^{-1}\boldsymbol{\tau}_h\left(X_j^c - x^c\right)$ $K_{h\lambda}(X_j,\, x)\mathbf{1}_j^y$, and $\overline{\mathbf{1}}_{z,y}(\varsigma_i) \equiv \mathbf{1}\{Z_i \leq z\} - F(z|X_i, y)$. Define

$$B_n \equiv \frac{(h!)^{1/2}}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\left[\mathbf{K}_r\left(\zeta_j, X_i\right)\overline{\mathbf{1}}_{Z_i,r}(\varsigma_j) - \mathbf{K}_s\left(\zeta_j, X_i\right)\overline{\mathbf{1}}_{Z_i,s}(\varsigma_j)\right]^2, \tag{4.9}$$

and

$$\begin{aligned}\sigma_n^2 \;\equiv\;& 2h!E_iE_j\left[\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\int\left\{\mathbf{K}_r(\zeta_i; x)\,\overline{\mathbf{1}}_{z,r}(\varsigma_i) - \mathbf{K}_s(\zeta_i, x)\,\overline{\mathbf{1}}_{z,s}(\varsigma_i)\right\}\left\{\mathbf{K}_r\left(\zeta_j; x\right)\overline{\mathbf{1}}_{z,r}(\varsigma_j)\right.\right.\\ &\left.\left. - \mathbf{K}_s(\zeta_i, x)\,\overline{\mathbf{1}}_{z,s}(\varsigma_i)\right\}F_\xi(d\xi)\right]^2 \end{aligned}\tag{4.10}$$

where E_i denote the expectation with respect to ξ_i. Let $\sigma_0^2 \equiv \lim_{n\to\infty}\sigma_n^2$.

Our first result says that after centering, $(h!)^{1/2}D_n$ is asymptotically normally distributed under H_0.

Theorem 4.1 *Suppose Assumptions A.1-A.5 hold. Then under* H_0, $(h!)^{1/2}D_n - B_n \xrightarrow{d} N\left(0, \sigma_0^2\right)$.

To implement the test, we need to consistently estimate B_n and σ_0^2. For this purpose, let $\widehat{\mathbf{1}}_{Z_i,y}(\varsigma_j) \equiv \mathbf{1}\{Z_j \leq Z_i\} - \widehat{F}(Z_i|X_j, y)$. Let $\widehat{\mathbf{K}}_r\left(\zeta_j; x\right) \equiv e_1'[\mathbf{S}_{nr}(x)]^{-1}\boldsymbol{\tau}_h\left(X_j^c - x^c\right)K_{h\lambda}(X_j, x)\,\mathbf{1}_j^y$. Let

$$\widehat{\alpha}_{ij,rs} \equiv \widehat{\mathbf{K}}_r\left(\zeta_j; X_i\right)\widehat{\mathbf{1}}_{Z_i,r}(\varsigma_j) - \widehat{\mathbf{K}}_s\left(\zeta_j; X_i\right)\widehat{\mathbf{1}}_{Z_i,s}(\varsigma_j), \text{ and } \widehat{\beta}_{ij,rs} \equiv \frac{1}{n}\sum_{l=1}^{n}\widehat{\alpha}_{li,rs}\widehat{\alpha}_{lj,rs}.$$

Define

$$\widehat{B}_n \equiv \frac{(h!)^{1/2}}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \widehat{\alpha}_{ij,rs}^2, \text{ and } \widehat{\sigma}_n^2 \equiv \frac{2h!}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \left[\sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \widehat{\beta}_{ij,rs} \right]^2.$$

We demonstrate in Theorem 4.2 below that $\widehat{B}_n - B_n = o_P(1)$ and $\widehat{\sigma}_n^2 - \sigma_0^2 = o_P(1)$. Then we can compare

$$T_n \equiv \left((h!)^{1/2} D_n - \widehat{B}_n \right) / \sqrt{\widehat{\sigma}_n^2} \tag{4.11}$$

to the one-sided critical value z_α, the upper α percentile from the $N(0,1)$ distribution. We reject the null at level α if $T_n > z_\alpha$.

To examine the asymptotic local power of the test, we consider the following sequence of Pitman local alternatives:

$$H_1(\gamma_n) : F(z|x,r) - F(z|x,s) = \gamma_n \delta_{n,rs}(\varsigma) \text{ for a.e. } \xi, \tag{4.12}$$

where $\gamma_n \to 0$ as $n \to \infty$ and $\delta_{n,rs}(\cdot)$ is a continuous function such that $\mu_0 \equiv \lim_{n\to\infty} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} E[\delta_{n,rs}(\varsigma_i)]^2 < \infty$. The following theorem establishes the local power of the test.

Theorem 4.2 *Suppose Assumptions A.1-A.5 hold. Then under $H_1(\gamma_n)$ with $\gamma_n = n^{-1/2}(h!)^{-1/4}$, $T_n \xrightarrow{d} N(\mu_0/\sigma_0, 1)$.*

Thus, the test has nontrivial power against Pitman local alternatives that converge to zero at rate $n^{-1/2}(h!)^{-1/4}$. The asymptotic local power function is given by $1 - \Phi(z_\alpha - \mu_0/\sigma_0)$, where Φ is the standard normal CDF.

The following theorem establishes the consistency of the test under the global alternative H_1 stated in (4.2).

Theorem 4.3 *Suppose Assumptions A.1-A.5 hold. Then under H_1, $n^{-1}(h!)^{-1/2} T_n = \mu_A/\sigma_0 + o_P(1)$, where $\mu_A \equiv \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} E[F(Z_i|X_i,r) - F(Z_i|X_i,s)]^2$, so that $P(T_n > c_n) \to 1$ under H_1 for any nonstochastic sequence $c_n = o\left(n(h!)^{1/2}\right)$.*

Remark 2. Alternatively, one can consider testing the conditional independence of Y and Z given X based upon the comparison of $F(z|x)$ with $F(z|x,y)$. In this case, the test statistic would be

$$\widetilde{D}_n \equiv \sum_{i=1}^{n} \left[\widetilde{F}(Z_i|X_i) - \widetilde{F}(Z_i|X_i,Y_i) \right]^2,$$

where $\widetilde{F}(z|x)$ and $\widetilde{F}(z|x,y)$ are local linear estimates of $F(z|x)$ and $F(z|x,y)$ by smoothing all discrete variables in X_i and (X_i, Y_i), respectively. After being suitably centered and rescaled, $\widetilde{D}_n$ can be shown to be asymptotically normally distributed. The key assumption for the asymptotic normality of $\widetilde{D}_n$ would require that the bandwidth (λ_y, say) used in smoothing the discrete variable Y_i tends to zero as $n \to \infty$. Nevertheless, under the null hypothesis of conditional independence, Y_i is an irrelevant variable in the prediction of Z_i or $\mathbf{1}(Z_i \leq z)$, implying that the optimal bandwidth for λ_y should tend to 1 as $n \to \infty$ (see Li and Racine (2007)). Thus this creates a dilemma for the choice of λ_y, making it extremely difficult to control the finite sample level of a test based upon $\widetilde{D}_n$. In contrast, when we construct our D_n test statistic, we obtain the estimate $\widehat{F}(Z_i|X_i,y)$ of $F(Z_i|X_i,y)$ for different values of y without smoothing the discrete variable Y_i (see (4.8)) and thus avoid the above dilemma.

5 Monte Carlo Simulations

In this section we conduct some Monte Carlo experiments to evaluate the finite sample performance of our test. We consider two data generating processes (DGPs):

DGP 1.

$$
\begin{aligned}
Y_i &= \mathbf{1}\left(\varepsilon_{Yi} \leq m_Y(X_i)\right), \\
Z_i &= \mathbf{1}\left(\varepsilon_{Zi} \leq m_Z(X_i)\right), \\
m_Y(X_i) &= \frac{X_{i1}^c - 0.5\left(X_{i1}^c\right)^2 + \phi\left(X_{i2}^c\right) - X_{i1}^c X_{i2}^c - 0.5 X_{i1}^c X_{i1}^d + 0.5 X_{i1}^d + 0.5 X_{i1}^d X_{i2}^d}{\sqrt{1 + X_{i1}^{c2} + X_{i2}^{c2}}}, \\
m_Z(X_i) &= \phi\left(X_{i1}^c\right) X_{i2}^c - X_{i1}^c - X_{i2}^c X_{i2}^d + 0.5 X_{i1}^d X_{i2}^d + \delta Y_i X_{i1}^c,
\end{aligned}
$$

where $X_i \equiv \left(X_{i1}^c, X_{i2}^c, X_{i1}^d, X_{i2}^d\right)'$, ϕ is the $N(0,1)$ PDF, X_{i1}^c is IID $U(0,4)$, X_{i2}^c is IID, computed as the sum of 48 independent random variables, each uniformly distributed on $[-0.25,\ 0.25]$, $P\left(X_{i1}^d = l\right) = 1/4$ for $l = 0,1,2,3$, $P\left(X_{i2}^d = l\right) = 1/5$ for $l = 0,1,2,3,4$, ε_{Y1} is IID $N(0,1)$, ε_{Zi} is IID $N(0,1)$, and all these variables are mutually independent. δ controls the degree of conditional dependence between Y_i and Z_i given X_i. Given X_i, Y_i and Z_i are conditionally independent when $\delta = 0$ and conditionally dependent otherwise.

DGP 2.

$$
\begin{aligned}
Y_i &= \mathbf{1}\left(\varepsilon_{Yi} \leq m_Y(X_i)\right), \\
Z_i &= m_Z(X_i) + s\ \varepsilon_{Zi},
\end{aligned}
$$

where $X_i \equiv \left(X_{i1}^c, X_{i2}^c, X_{i1}^d, X_{i2}^d\right)'$, ε_{Yi} and ε_{Zi} are generated as in DGP 1, m_Y and m_Z are as defined in DGP1, and s is taken to ensure the signal-noise ratio in the equation for Z_i to be 1 across all simulations.

Clearly, DGP 1 generates binary Y_i and Z_i variables whereas DGP 2 generates binary Y_i and continuous Z_i. In both DGPs the X_i vector includes two continuous variables, X_{i1}^c and X_{i2}^c, and two discrete variables, X_{i1}^d and X_{i2}^d. Note that our test is based on local linear regressions, which typically requires compactly supported conditioning variables. This motivates the otherwise awkward way we generate X_{i2}^c in DGPs 1-2. According to the central limit theorem, we can treat X_{i2}^c as being nearly standard normal random variables but with compact support $[-12, 12]$.

Notice that the two discrete variables in X_i partition the data into $4 \times 5 = 20$ cells. In conjunction with the 2 categories of dummy Y_i, this will partition the data into $20 \times 2 = 40$ cells if we adopt the conventional *nonparametric frequency approach* to do the estimation and testing. If the number of observations n is small, say, 100, each cell has a tiny amount of observations on average and some empty cells in practice, this will make the estimation of the CDF $F(z|x, y)$ extremely difficult. A nonparametric-frequency-based test should not be expected to perform well in terms of both level and power. Even with nonparametric smoothing over the discrete variables in X_i as advocated by our test, the problem continues to be hard but less severe.

To construct the test statistic, we need to choose both kernel and bandwidth. We choose the product of Gaussian kernel for the two continuous regressors: $q(x) = (2\pi)^{-1/2} \exp\left(-x^2/2\right)$. Since there is no data-driven procedure to choose the bandwidths $h = (h_1, h_2)'$ and $\lambda = (\lambda_1, \lambda_2)'$ for our testing problem, we choose them according to the rule of thumb:

$$h_l = \gamma s_{X_l^c} n^{-1/4.5},\ \lambda_l = \gamma n^{-2/4.5} \text{for } l = 1, 2, \tag{5.1}$$

where $s_{X_l^c}$ is the sample standard deviation of X_{il}^c and γ is a fixed constant. We study the behavior of our test with different choices of γ in order to examine the sensitivity of our test to the bandwidth sequence. Robinson (1991, p.448) proposes very similar devices. Note that these choices for h and λ and the kernel function meet the requirements for our test. Through a preliminary simulation study, we find our bootstrap-based test is not sensitive to the choice of γ when we take $\gamma \in [0.5, 2]$. So we fix $\gamma = 1$ for our simulation results.

It is well known that the asymptotic normal distribution typically cannot approximate the finite sample distribution of many nonparametric test statistics. This is especially true for our test when we have discrete conditioning variables in X_i with reasonably large number of categories. So we suggest using a bootstrap method to obtain the bootstrap p-values. Here, we generate the bootstrap data $\{(X_i^*, Y_i^*, Z_i^*)\}_{i=1}^n$ based on the following local bootstrap procedure:

1. Set $(X_i^*, Y_i^*) = (X_i, Y_i)$ for each $i \in \{1, \cdots, n\}$.

2. For $i = 1, \cdots n$, given X_i^*, draw Z_i^* from the following local constant nonparametric estimate of $F(z|X_i^*)$:

$$\widetilde{F}_{\widetilde{h}\widetilde{\lambda}}(z|X_i^*) = \frac{\sum_{j=1}^n K_{\widetilde{h}\widetilde{\lambda}}(X_j, X_i^*) \mathbf{1}(Z_j \leq z)}{\sum_{j=1}^n K_{\widetilde{h}\widetilde{\lambda}}(X_j, X_i^*)} \tag{5.2}$$

   where $\widetilde{h}$ and $\widetilde{\lambda}$ are the bandwidth used in the estimation of $F(z|X_i^*)$.

3. Compute the bootstrap test statistic T_n^* in the same way as T_n by using $\{(X_i^*, Y_i^*, Z_i^*)\}_{i=1}^n$ instead.

4. Repeat steps 1-3 B times to obtain B bootstrap test statistic $\left\{T_{nj}^*\right\}_{j=1}^B$. Calculate the bootstrap p-values $p^* \equiv B^{-1} \sum_{j=1}^B \mathbf{1}\left(T_{nj}^* \geq T_n\right)$ and reject the null hypothesis of conditional independence if $p^*$is smaller than the prescribed level of significance.

The above procedure is coined as the local bootstrap procedure by Paparoditis and Politis (2000) who also explain how to generate the bootstrap observations computationally. It works no matter whether Z_i is discrete or continuous. In the case where Z_i is continuous, we can also generate a smooth version of Z_i^* through $Z_i^{**} = Z_i^* + b\eta_i$, where $b \equiv b(n) \to 0$ as $n \to \infty$, and η_i is drawn from $N(0, 1)$. In our simulations and applications, we generate Z_i^* and Z_i^{**} for the case where Z_i is discrete and continuous, respectively. When Z_i is continuous, we set $b = s_Z n^{-1/6}$ with s_Z being the sample standard deviation of Z_i. Our simulations indicate that the choice of b plays little role in the performance of our test. For simplicity, we set $\widetilde{h} = h$ and $\widetilde{\lambda} = \lambda$.

Table 1 reports the empirical rejection frequencies of our test at 1%, 5%, and 10% nominal levels for DGPs 1-2. Also reported in the table is a variant of our test based on the idea of nonparametric frequency, which is obtained by setting the smoothing parameters for the discrete variables in X_i to be 0 in the calculation of our test statistic. To save on computational time, we use 250 replications for each sample size n and 100 bootstrap resamples in each replication. We summarize some important findings from Table 1.

First, the level of our nonparametric smoothing test is reasonably well behaved despite the fact that it tends to be oversized when n is small and the average number of observations

Table 1: Finite sample rejection frequency for DGPs 1-2

DGP	Sample size n	δ	Our test $h_l = s_{X_l^c} n^{-1/4.5},\ \lambda_l = n^{-2/4.5}$ 1%	5%	10%	Nonparametric frequency approach $h_l = s_{X_l^c} n^{-1/4.5},\ \lambda_l = 0$ 1%	5%	10%
1	200	0	0.036	0.092	0.144	0.100	0.224	0.288
		1	0.420	0.592	0.644	0.200	0.268	0.348
		2	0.900	0.920	0.924	0.472	0.620	0.656
	400	0	0.024	0.068	0.104	0.088	0.216	0.296
		1	0.628	0.720	0.764	0.156	0.312	0.384
		2	0.904	0.924	0.932	0.512	0.648	0.728
	800	0	0.016	0.040	0.068	0.160	0.348	0.448
		1	0.840	0.880	0.884	0.204	0.396	0.492
		2	0.968	0.968	0.968	0.628	0.752	0.836
2	200	0	0.020	0.068	0.124	0.028	0.080	0.176
		1	0.064	0.160	0.268	0.068	0.172	0.232
		2	0.176	0.300	0.448	0.088	0.240	0.312
	400	0	0.004	0.016	0.072	0.032	0.132	0.196
		1	0.080	0.192	0.288	0.068	0.172	0.248
		2	0.268	0.504	0.632	0.100	0.196	0.328
	800	0	0.000	0.032	0.056	0.020	0.128	0.216
		1	0.168	0.304	0.408	0.100	0.276	0.360
		2	0.600	0.768	0.812	0.136	0.296	0.380

per cell is small. In the case where $n = 200$, $X_i^d \equiv (X_{i1}^d, X_{i2}^d)'$ and Y_i partition the 200 observations into 40 cells so that each cell contains only 5 observations on average. Given the two conditioning variables X_{i1}^c and X_{i2}^c, one cannot expect the conditional CDF for each cell values of X_i^d and Y_i to be well estimated no matter whether we choose to smooth X_i^d or not. This definitely has some adverse effect on the performance of our test. Despite this, our nonparametric smoothing test seems to perform well even if n is small and the average number of observations per cell is small. As n and the average number of observations per cell double, the levels of our test tend to be improved and get close to the nominal levels.

Second, our test has power to detect deviations from conditional independence no matter whether Z_i is discrete or continuous. In DGP 1 when δ changes from 0 to 1 (resp. 2) so that Y_i becomes to affect Z_i conditional on X_i, the unconditional probability for Z_i to take value 1 increases from 0.38 to 0.52 (resp. 0.60). Our nonparametric smoothing test can detect such changes very well even for small n. As n doubles, the above changes of unconditional probabilities remain the same as we change δ, but the power of our test increases. In DGP 2, Z_i is continuously valued. The power performance does not appear to be as well as the case of DGP 1 because we normalize the error terms in the equation for Z_i to ensure the signal-noise ratio to be 1 across different values of δ. If we set $s = 1$ in the equation for Z_i and allow the signal become stronger as δ increases, we can observe significant improvement of the power performance of our test.

Third, in terms of both size and power, our smoothing nonparametric test significantly dominates the nonparametric-frequency-based test. The latter test tends to be oversized for both DGPs and all sample sizes under investigation. Despite its oversize, as expected, the latter test is much less powerful in detecting deviations from the null of conditional independence than our nonparametric smoothing test.

6 Empirical Applications

In this section we apply the nonparametric test to an automobile insurance data set.[10] We first briefly introduce the automobile insurance market in France where our data set stems from, then discuss configurations of the data set and present our empirical findings. Noting that the design of automobile insurance is relatively similar in most countries, so we believe that our methodology is broadly applicable.

[10] Despite the scarcity of insurance data sets the car insurance has been analyzed for different countries amongst others by Chiappori and Salanié (1997, 2000), Richaudeau (1999), Cohen (2005), Saito (2006) and Kim et. al. (2009).

6.1 Principles of the Automobile Insurance in France

In France, like in many other countries, all cars must be insured at the "responsabilité civile" (RC) level. This is a liability insurance that covers damage inflicted to other drivers and their cars. Moreover, insurance companies offer additional non-compulsory coverage. The most common one is called "assurance tous risques" (TR), which also covers damage to the insured car or the driver in the case of an accident at which he or she is at fault. The insurees can choose from different comprehensive insurance contracts which vary in the value of the deductible (fixed or proportional).

A special feature of the car insurance is the so called "bonus/ malus", a uniform experience rating system. At any date/year t, the premium is defined as the product of a basis amount and a "bonus" coefficient. The basic amount can be defined freely by the insurance companies according to their risk classification but cannot be related to past experience. The past experience is captured by the so called "bonus/ malus" coefficient whose evolution is strictly regulated. Suppose, the bonus coefficient is b_t at the beginning of the tth period. Then the occurrence of an accident during the period leads to an increase of 25 percent at the end of the period (i.e., $b_{t+1} = 1.25b_t$), whereas an accident-free year implies a reduction of 5 percent at the end (i.e., $b_{t+1} = 0.95b_t$). Additionally, several special rules are applied, which include the permission to overcharge contracts held by young drivers. But the surcharge is limited to 140 percent of the basis rate and is forced to decrease by half every year in which the insuree has not had an accident.

The basis amount of the premium is calculated according to different risk classes. Due to variables like age, sex, profession, area, etc., the insurees are divided into different risk classes which should reflect their accident probabilities, and the premium to be paid is then determined.

6.2 Configurations of the Data Set

We use a data set of the French federation of insurers[11] (FFSA) which conducted in 1990 a survey of its members. This data set was also used in Chiappori and Salanié (1997, 2000). With a sampling rate of 1/20 the data set consists of 41 variables on $1,120,000$ contracts and 25 variables on $120,000$ accidents for the year 1989. For each driver all variables which are used by insurance companies for pricing their contracts - age of the driver, sex, profession of the driver, year of drivers license, age of the car, type of the car, use of the car, and area - plus the characteristics of the contract and the characteristics of the accident, if occurred, are

[11] The FFSA comprehends 21 companies that together have 70 percent market share of the French automobile insurance market.

available. We restrict our analysis to all "young"[12] drivers who obtained their driver license in 1988. This reduces the sample size to $n = 6,333$.

As Chiappori and Salanié (2000) argue, focusing on young drivers has two major advantages. In a subsample of young drivers the driving experience is much more homogeneous than that in the total population in which groups of different experiences are pooled. Therefore the heteroskedasticity problem is mitigated and less severe. The concentration on young drivers also avoids the problems associated with the experience rating and the resulting bias. The past driving history is usually observed by the insurance companies. The past driving records are highly informative on probabilities of accident and used for pricing. The bonus coefficient is a very excellent proxy for this variable. However, the introduction of this variable is quite delicate because of its endogeneity. This problem can be circumvented either by using panel data[13] or by using only data on beginners. We pursue the second approach and concentrate on novice drivers.

One important issue in testing for asymmetric information is the distinction between accidents and claims. The data set of insurance companies comprise claims. But whether an accident - once it has occurred - is declared to the insurance company and becomes a claim depends on the decision of the insuree. This decision is mainly determined by the nature of the contract. For example, accidents whose damage is below the deductible or is not covered are usually not declared. Therefore one might expect a positive correlation between the type of contract (coverage) and the probability of a claim - even in the absence of ex ante moral hazard.[14] One strategy to handle this problem is to discard all accidents in which only one automobile was involved. Whenever two cars are involved, a declaration is nearly inevitable.

To make the results comparable with those of Chiappori and Salanié (2000) and to check for robustness we examine several different configurations of the data set. Let X_i denote the set of exogenous control variables for individual i. Let $Y_i = 0$ if individual i buys only the minimum legal coverage (a RC contract) and 1 if individual i buys any form of comprehensive coverage (a TR contract). First we consider discrete Z_i where $Z_i = 1$ if i has at least one accident in which he or she is judged to be at fault and 0 otherwise (no accident occurred or i was not at fault). Then we consider the case where Z_i is continuous and defined by the total payments caused by the insuree, which is also included in the data set.

For the random variables in X_i, we consider three configurations. In Configuration I we include the following control variables in X_i : sex (2), make of car (8), performance of the

[12] "Young" refers not to the actual age but to the driving experience.
[13] For a detailed discussion see Chiappori and Heckmann (1999).
[14] The phenomenon that accidents that are not covered are not declared is sometimes called "ex post moral hazard".

car (6), type of use (4), type of area (5), profession of the driver (8), region (10), age of the driver, and age of the car, where numbers in brackets indicate the number of categories for the corresponding discrete variables, and variables without numbers indicate they are continuous variables. These control variables are similar to those used by Chiappori and Salanié (2000) for their probit-model- or χ^2-based tests except that we do not transform the age of the car and that of the driver to discrete variables.

Our nonparametric test requires that the number of observations per cell should not be too small. So we also consider another two configurations for X_i. In Configuration II we omit the variable, make of the car, which describes the home country of the manufacturer of the car. We think that the most important part of the information concerning an automobile can be captured by the performance of the car, so that the omission of this variable should have no significant influence on the results. For example, the accident probability of an Italian and a French compact car should not differ significantly, all other things being equal. Additionally, we reduce the number of categories for some discrete variables according to Column 3 in Table 2. Again, we argue that merging categories which are nearly identical or closely related does not bias the results.

In Configuration III we use only two categories for each of the seven discrete variables in X_i. As surveyed above, Salanié and Chiappori (2000) also conduct nonparametric tests where they code all control variables as binary and apply a χ^2 independence test to each cell, and then aggregate the resulting test statistics in three different ways. Our third configuration enables a direct comparison of our nonparametric test with their nonparametric tests.

Configurations IV - VI correspond to Configurations I - III, respectively. In the settings IV - VI we only replace the discrete dummy variable Z_i by its continuous counterpart, i.e., by the total payments caused through accidents by the insuree to the insurance company. In all configurations, we treat the age of the car and the age of the driver as continuous variables. See Table 2 for a summary of these configurations.

6.3 Empirical Results

In this subsection we apply the nonparametric test to the data set introduced in the above subsection. Table 3 reports the bootstrap p-values for our nonparametric test under various configurations of the data set. Given the large sample size ($n = 6,333$) and the need of bootstrap, the computational burden for our bootstrap-based nonparametric test is very heavy. Simulations for smaller sample sizes with different choices of the number of bootstrap replications ($B = 100$, 200, 300) indicate that our testing results are insensitive to the choice of B. So we only set $B = 100$ for our applications. Also due to the large sample size and the

Table 2: An overview of the data configurations

Variables\Configurations	I	II	III	IV	V	VI
Y_i	2	2	2	2	2	2
Z_i	2	2	2	X	X	X
sex	2	2	2	2	2	2
make of car	8	-	2	8	-	2
performance of car	6	6	2	6	6	2
type of use	4	3	2	4	3	2
type of area	5	2	2	5	2	2
profession of driver	9	5	2	9	5	2
region	10	5	2	10	5	2
age of driver	X	X	X	X	X	X
age of car	X	X	X	X	X	X

Note: Integers denote the number of categories for the corresponding discrete variables. An "X" in the table denotes the corresponding variable is a continuous variable.

Table 3: Bootstrap p values for our nonparametric test under various configurations

γ\Configurations	I	II	III	IV	V	VI
$\gamma = 0.75$	0.82	0.54	0.24	1.00	1.00	1.00
$\gamma = 1$	0.79	0.62	0.13	1.00	1.00	1.00
$\gamma = 1.25$	0.77	0.66	0.14	1.00	1.00	1.00
$\gamma = 1.5$	0.76	0.72	0.21	1.00	1.00	1.00

large number of control variables, it is difficult to use least squares cross validation method to choose data-driven bandwidths to conduct our nonparametric test. For this reason we adopt the rule of thumb to choose the bandwidths: $h_l = \gamma s_{X_l^c} n^{-1/4.5}$ and $\lambda_s = \gamma n^{-2/4.5}$ for the continuous and discrete control variables, respectively. To check the sensitivity of the test to the choice of bandwidth, we consider four values of γ : 0.75, 1, 1.25, and 1.5. These choices of bandwidths fulfill the requirements of Assumption A.5.

We summarize some important findings from Table 3. First, it indicates that in all cases we fail to reject the null hypothesis of absence of asymmetric information at the 10% significance level. This means that the knowledge of the choice of the contract does not contain information for predicting the probability of an accident or the other way round, knowing the number of accidents (discrete) or the caused damages (continuous) is of no value for predicting the chosen contract. Therefore our test affirms the Chiappori and Salanié's (2000) findings that there is no evidence of asymmetric information in the market for automobile insurance in France. The results are very robust to different configurations of data and choices of bandwidth. Second, Table 3 reveals that an aggregation of the categories of the discrete control variables leads to a decrease of the p-values so that a reduction of information might disguise asymmetric information. Therefore, a (non-)parametric test that relies on highly aggregated information might yield misleading or wrong conclusions. Third, Table 3 also reveals that using the payments of the insurance companies instead of the number of accidents leads to a strengthening of the absence of asymmetric information. Again, a reduction of information might lead to wrong test conclusions.

Recently Kim et al. (2009) have argued that the absence of asymmetric information in most empirical studies might be due to the dichotomous measurement approach that induces the excessive bundling of contracts with different deductibles. In reality the insurees can choose between several deductibles referring to different fields of coverage. But most studies aggregate this choice opportunities to a binary choice between "compulsory" coverage and "additional" coverage so that the choice variable Y_i becomes binary. Kim et al. (2009) claim that excessive bundling in coverage measurements might disguise the existence of asymmetric information. So they apply a multinomial measurement approach, which is parametric in nature, and demonstrate the evidence of asymmetric information in their data set obtained from a major automobile insurance company in Korea.

Since our data set also contains the exact level of the chosen deductible, we can investigate this hypothesis as our test is fully applicable to this problem. A very small proportion of the contracts has proportional deductibles which are dropped for this analysis. Therefore the sample size decreases to $n = 6,219$. We divide the chosen deductible into three ($0 - 100$, $101 - 1500$, and > 1500) groups. The results are reported in Table 4 for different data configurations introduced above. In comparison to the settings defined in Table 2, the choice variable Y_i now has three categories, but everything else remains unchanged in the data set. Clearly, Table 4 confirms the absence of asymmetric information in the data. We also tried a finer division for the deductible so that Y_i has more categories. In all cases, our results are robust in that they all confirm the absence of asymmetric information in the data. Intuitively speaking, if there is no asymmetric information in the most important choice between compulsory and comprehensive insurance, one should not expect asymmetric information in the minor decision of the exact deductible when the money at stack is not so high.

Table 4: Bootstrap p values for our nonparametric test when the choice variable has three deductible levels

$\gamma \backslash$Configurations	I	II	III
$\gamma = 0.75$	0.71	0.73	0.73
$\gamma = 1$	0.67	0.71	0.88
$\gamma = 1.25$	0.69	0.64	0.68
$\gamma = 1.5$	0.64	0.77	0.75

7 Concluding Remarks

We propose a new nonparametric test for asymmetric information in this paper and apply it to a French automobile insurance data set. Our main conclusion is that we cannot detect asymmetric information in the data set despite different configurations of the control variables and different choices of bandwidth parameters. Our nonparametric test does not require specification of any functional or distributional form among the sets of variables of interest and it is not subject to any misspecification problem given the right choice of control variables. We also show that excessive bundling does not necessarily result in a disguise of asymmetric information. Both in the case of the binary choice between "compulsory" coverage and "additional" coverage and in the case of several deductibles (three and more groups) we confirm the absence of asymmetric information. Our results are also very strong in contrast to Kim et al. (2009).

Since nearly all other classes of insurance, such as the legal protection insurance, private health insurance, and disability insurance, are structured in the same way as the auto insurance, applications to data sets in these subfields are immediate and might help to gain new insights. Moreover, our test can be applied to more general settings, either to testing for asymmetric information in other fields or more generally, to testing the general hypothesis of conditional independence.

A Mathematical Appendix

Let $\Delta_{j,y}(x,z) \equiv F(z|X_j,y) - F(z|x,y) - \sum_{s=1}^{p_c} \left(\partial F\left(z|x^c,x^d,y\right)/\partial x_s^c\right)\left(X_{js}^c - x_s^c\right)$, $\mathbf{V}_{ny}(\varsigma) \equiv \frac{1}{n}\sum_{j=1}^n K_{h\lambda}(X_j,x)\,\mathbf{1}_j^y \boldsymbol{\tau}_h(X_j^c - x^c)\overline{\mathbf{1}}_{z,y}(\varsigma_j)$, and $\mathbf{B}_{ny}(\varsigma) \equiv \frac{1}{n}\sum_{j=1}^n K_{h\lambda}(X_j,x)\,\mathbf{1}_j^y \boldsymbol{\tau}_h(X_j^c - x^c)\Delta_{j,y}(x,z)$. Let $\overline{\mathbf{S}}_y(x) \equiv E[\mathbf{S}_{ny}(x)]$ and $\overline{\mathbf{B}}_y(\varsigma) \equiv E[\mathbf{B}_{ny}(\varsigma)]$. The following lemma establishes the uniform consistency of $\widehat{F}(z|x,y)$.

Lemma A.1 *Suppose Assumptions A.1-A.5 hold. Then uniformly in* $\xi \equiv (x',y,z)'$ *we have:* $\widehat{F}(z|x,y) - F(z|x,y) = e_1'[\overline{\mathbf{S}}_y(x)]^{-1}[\mathbf{V}_{ny}(\varsigma) + \overline{\mathbf{B}}_y(\varsigma)] + O_P(\nu_n^2 + \nu_n(\|h\|^2 + \|\lambda\|)) = O_P(\nu_n + \|h\|^2 + \|\lambda\|)$, *where* $\nu_n \equiv n^{-1/2}(h!)^{-1/2}\sqrt{\log n}$.

Proof. Since $[\mathbf{S}_{ny}(x)]^{-1}\,\mathbf{S}_{ny}(x) = I_{p_c+1}$ where I_{p_c+1} is a $(p_c+1)\times(p_c+1)$ identity matrix, we obtain the following standard bias and variance decomposition:

$$\widehat{F}(z|x,y) - F(z|x,y) = e_1'[\mathbf{S}_{ny}(x)]^{-1}\mathbf{V}_{ny}(\varsigma) + e_1'[\mathbf{S}_{ny}(x)]^{-1}\mathbf{B}_{ny}(\varsigma), \tag{A.1}$$

where e_1' is the first row of I_{p_c+1}. By Theorems 2 and 4 in Masry (1996) with little modification to account for discrete regressors,[15]

$$\mathbf{S}_{ny}(x) = \overline{\mathbf{S}}_y(x) + O_P(\nu_n),\, \mathbf{V}_{ny}(\varsigma) = O_P(\nu_n), \text{ and } \mathbf{B}_{ny}(\varsigma) - \overline{\mathbf{B}}_y(\varsigma) = O_P(\nu_n(\|h\|^2 + \|\lambda\|)),$$

where the probability orders hold uniformly in $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. By the Slutsky lemma,

$$[\mathbf{S}_{ny}(x)]^{-1} = \left\{\overline{\mathbf{S}}_y(x) + \left[\mathbf{S}_{ny}(x) - \overline{\mathbf{S}}_y(x)\right]\right\}^{-1} = [\overline{\mathbf{S}}_y(x)]^{-1} + O_P(\nu_n). \tag{A.2}$$

By the same argument as used in the proof of Theorem 4.1 of Boente and Fraiman (1991), we can show that $\mathbf{V}_{ny}(\varsigma) = O_P(\nu_n)$ uniformly in ς under Assumption A.3. It follows that $\widehat{F}(z|x,y) - F(z|x,y) = e_1'\{[\overline{\mathbf{S}}_y(x)]^{-1} + O_P(\nu_n)\}\{\mathbf{V}_{ny}(\varsigma) + [\overline{\mathbf{B}}_y(\varsigma) + O_P(\nu_n(\|h\|^2 + \|\lambda\|)]\}$ $= e_1'[\overline{\mathbf{S}}_y(x)]^{-1}[\mathbf{V}_{ny}(\varsigma) + \overline{\mathbf{B}}_y(\varsigma)] + O_P(\nu_n^2 + \nu_n(\|h\|^2 + \|\lambda\|)) = O_P(\nu_n + \|h\|^2 + \|\lambda\|)$. ■

Proof of Theorems 4.1 and 4.2

We only prove Theorem 4.2, as the proof of Theorem 4.1 is a special case.

[15] The compact support of the kernel function in Masry (1996) can be easily relaxed, following the line of proof in Hansen (2008, Theorem 4). Masry (1996) only allows continuous regressors, which can also be extended to the case of mixed regressors. Since X_i^d and Y_i only take finite number of possible values, they have no impact on the uniform probability order.

First, we decompose $(h!)^{1/2} D_n$ as follows:

$$\begin{aligned}
(h!)^{1/2} D_n &= (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \left[\widehat{F}(Z_i|X_i,r) - \widehat{F}(Z_i|X_i,s)\right]^2 \\
&= (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \Big\{ \left[F(Z_i|X_i,r) - F(Z_i|X_i,s)\right]^2 \\
&\quad + \left[\widehat{F}(Z_i|X_i,r) - F(Z_i|X_i,r) - \widehat{F}(Z_i|X_i,s) + F(Z_i|X_i,s)\right]^2 \\
&\quad + 2\left[F(Z_i|X_i,r) - F(Z_i|X_i,s)\right] \\
&\quad \times \left[\widehat{F}(Z_i|X_i,r) - F(Z_i|X_i,r) - \widehat{F}(Z_i|X_i,s) + F(Z_i|X_i,s)\right]\Big\} \\
&\equiv D_{n1} + D_{n2} + 2D_{n3}.
\end{aligned}$$

Under $H_1(n^{-1/2}(h!)^{-1/4})$, we prove the theorem by showing that (i) $D_{n1} \xrightarrow{P} \mu_0$, (ii) $D_{n2} - B_n \xrightarrow{d} N\left(0, \sigma_0^2\right)$, (iii) $D_{n3} = o_P(1)$, (iv) $\widehat{B}_n = B_n + o_P(1)$, and (v) $\widehat{\sigma}_n^2 = \sigma_0^2 + o_P(1)$. For (i), $D_{n1} = n^{-1} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \delta_{n,rs}(\varsigma_i)^2 = \mu_0 + o_P(1)$ under $H_1(n^{-1/2}(h!)^{-1/4})$. It remains to show (ii)-(iv).

To show (ii), we first apply Lemma A.1 to obtain

$$\begin{aligned}
D_{n2} &= (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \left[\widehat{F}(Z_i|X_i,r) - F(Z_i|X_i,r) - \widehat{F}(Z_i|X_i,s) + F(Z_i|X_i,s)\right]^2 \\
&= (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \Big[e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1}\mathbf{V}_{nr}(\varsigma_i) - e_1'[\overline{\mathbf{S}}_s(X_i)]^{-1}\mathbf{V}_{ns}(\varsigma_i) \\
&\quad + e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1}\overline{\mathbf{B}}_r(\varsigma_i) - e_1'[\overline{\mathbf{S}}_s(X_i)]^{-1}\overline{\mathbf{B}}_s(\varsigma_i) + O_P(\nu_n^2 + \nu_n(\|h\|^2 + \|\lambda\|))\Big]^2 \\
&= (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \left[e_1'\left\{[\overline{\mathbf{S}}_r(X_i)]^{-1}\mathbf{V}_{nr}(\varsigma_i) - [\overline{\mathbf{S}}_s(X_i)]^{-1}\mathbf{V}_{ns}(\varsigma_i)\right\}\right]^2 \\
&\quad + 2(h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} e_1'\left\{[\overline{\mathbf{S}}_r(X_i)]^{-1}\mathbf{V}_{nr}(\varsigma_i) - [\overline{\mathbf{S}}_s(X_i)]^{-1}\mathbf{V}_{ns}(\varsigma_i)\right\} \\
&\qquad \times e_1'\left\{[\overline{\mathbf{S}}_r(X_i)]^{-1}\overline{\mathbf{B}}_r(\varsigma_i) - [\overline{\mathbf{S}}_s(X_i)]^{-1}\overline{\mathbf{B}}_s(\varsigma_i)\right\} \\
&\quad + (h!)^{1/2} \sum_{i=1}^{n} \left[e_1'\left\{[\overline{\mathbf{S}}_r(X_i)]^{-1}\overline{\mathbf{B}}_r(\varsigma_i) - [\overline{\mathbf{S}}_s(X_i)]^{-1}\overline{\mathbf{B}}_s(\varsigma_i)\right\}\right]^2 \\
&\quad + n(h!)^{1/2} O_P\left(\nu_n^2 + \nu_n(\|h\|^2 + \|\lambda\|)\right) O_P\left(\nu_n + \|h\|^2 + \|\lambda\|\right) \\
&\equiv D_{n21} + 2D_{n22} + D_{n23} + o_P(1) \qquad \text{(A.3)}
\end{aligned}$$

where the definitions of D_{n21}, D_{n22}, and D_{n23} are self-evident. Using the notation defined above eq. (4.9), we have $D_{n21} = \frac{(h!)^{1/2}}{(n-1)^2} \sum_{i=1}^n \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} [\sum_{j=1}^n \varphi_{rs}\left(\xi_i, \xi_j\right)]^2$, where $\varphi_{rs}\left(\xi_i, \xi_j\right) \equiv \mathbf{K}_r\left(\zeta_j; X_i\right) \overline{\mathbf{1}}_{Z_i,r}\left(\varsigma_j\right) - \mathbf{K}_s\left(\zeta_j; X_i\right) \overline{\mathbf{1}}_{Z_i,s}\left(\varsigma_j\right)$. Decompose D_{n21} as follows

$$\begin{aligned}
D_{n21} &= \frac{(h!)^{1/2}}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \varphi_{rs}\left(\xi_i, \xi_j\right) \varphi_{rs}\left(\xi_i, \xi_k\right) \\
&= \frac{(h!)^{1/2}}{n^2} \sum_{i=1}^n \sum_{j\neq i}^n \sum_{k\neq i,j}^n \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \varphi_{rs}\left(\xi_i, \xi_j\right) \varphi_{rs}\left(\xi_i, \xi_k\right) \\
&\quad + \frac{(h!)^{1/2}}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \varphi_{rs}\left(\xi_i, \xi_j\right)^2 \\
&\quad + \frac{2\,(h!)^{1/2}}{n^2} \sum_{i=1}^n \sum_{j\neq i}^n \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \varphi_{rs}\left(\xi_i, \xi_j\right) \varphi_{rs}\left(\xi_i, \xi_i\right) \\
&\equiv V_n + B_n + R_n, \text{ say.}
\end{aligned} \tag{A.4}$$

Let $\overline{\varphi}_{rs}\left(\xi_i, \xi_j, \xi_k\right) \equiv [\varphi_{rs}\left(\xi_i, \xi_j\right) \varphi_{rs}\left(\xi_i, \xi_k\right) + \varphi_{rs}\left(\xi_j, \xi_i\right) \varphi_{rs}\left(\xi_j, \xi_k\right) + \varphi_{rs}\left(\xi_k, \xi_i\right) \varphi_{rs}\left(\xi_k, \xi_j\right)]/3$. Then

$$V_n = \frac{6\,(h!)^{1/2}}{n^2} \sum_{1\leq i<j<k\leq n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \overline{\varphi}_{rs}\left(\xi_i, \xi_j, \xi_k\right) = \frac{(n-1)\,(n-2)}{n} \overline{V}_n,$$

where $\overline{V}_n \equiv \frac{6(h!)^{1/2}}{n(n-1)(n-2)} \sum_{1\leq i<j<k\leq n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \overline{\varphi}_{rs}\left(\xi_i, \xi_j, \xi_k\right)$. Note that for all $i \neq j \neq k$, $\theta \equiv E\left[\overline{\varphi}_{rs}\left(\xi_i, \xi_j, \xi_k\right)\right] = 0$, $\overline{\varphi}_{rs,1}\left(a\right) \equiv E\left[\overline{\varphi}_{rs}\left(a, \xi_j, \xi_k\right)\right] = 0$, and $\overline{\varphi}_{rs,2}\left(a, \widetilde{a}\right) \equiv E\left[\overline{\varphi}_{rs}\left(a, \widetilde{a}, \xi_k\right)\right] = \frac{1}{3} E[\varphi_{rs}\left(\xi_k, a\right)\ \varphi_{rs}\left(\xi_k, \widetilde{a}\right)]$. Let $\overline{\varphi}_{rs,3}\left(a, \widetilde{a}, \overline{a}\right) \equiv \overline{\varphi}_{rs}\left(a, \widetilde{a}, a\right) - \overline{\varphi}_{rs,2}\left(a, \widetilde{a}\right) - \overline{\varphi}_{rs,2}\left(a, \overline{a}\right) - \overline{\varphi}_{rs,2}\left(\widetilde{a}, a\right)$. By the Hoeffding decomposition,

$$\overline{V}_n = 3H_n^{(2)} + H_n^{(3)},$$

where $H_n^{(2)} \equiv \frac{2(h!)^{1/2}}{n(n-1)} \sum_{1\leq i<j\leq n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \overline{\varphi}_{rs,2}\left(\xi_i, \xi_j\right)$ and $H_n^{(3)} \equiv \frac{6(h!)^{1/2}}{n(n-1)(n-2)} \sum_{1\leq i<j<k\leq n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \overline{\varphi}_{rs,3}(\xi_i, \xi_j, \xi_k)$. Noting that $E\left[\overline{\varphi}_{rs,3}\left(a, \widetilde{a}, \xi_i\right)\right] = 0$ and that $\overline{\varphi}_{rs,3}$ is symmetric in its arguments by construction, it is straightforward to show that $E[H_n^{(3)}] = 0$ and $E[H_n^{(3)}]^2 = O(n^{-3}\,(h!)^{-1})$. Hence, $H_n^{(3)} = O_P(n^{-3/2}\,(h!)^{-1/2}) = o_P\left(n^{-1}\right)$ by the Chebyshev inequality. It follows that $V_n = \frac{n(n-2)}{n-1} \overline{V}_n = \{1 + o\left(1\right)\} \mathcal{H}_n + o_P\left(1\right)$, where

$$\mathcal{H}_n \equiv \frac{2\,(h!)^{1/2}}{n} \sum_{1\leq i\leq j\leq n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} 3\overline{\varphi}_{rs,2}\left(\xi_i, \xi_j\right)$$

$$= \frac{2(h!)^{1/2}}{n} \sum_{1 \le i < j \le n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \int \varphi_{rs}(a, \xi_i) \varphi_{rs}(a, \xi_j) F_\xi(da).$$

As $\mathcal{H}_n$ is a second order degenerate U-statistic, it is straightforward but tedious to verify that all the conditions of Theorem 1 of Hall (1984) are satisfied, implying that a central limit theorem applies to $\mathcal{H}_n$: $\mathcal{H}_n \xrightarrow{d} N(0, \sigma_0^2)$, where the asymptotic variance of $\mathcal{H}_n$ is given by $\sigma_0^2 \equiv \lim_{n\to\infty} \sigma_n^2$ and $\sigma_n^2 \equiv 2h! E_i E_j [\sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \int \varphi_{rs}(\xi, \xi_i) \varphi_{rs}(\xi, \xi_j) F_\xi(d\xi)]^2 = 2h! E_i E_j [\sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \int [\mathbf{K}_r(\zeta_i; x) \overline{\mathbf{1}}_{z,r}(\varsigma_i) - \mathbf{K}_s(\zeta_i; x) \overline{\mathbf{1}}_{z,s}(\varsigma_i)] [\mathbf{K}_r(\zeta_j; x) \overline{\mathbf{1}}_{z,r}(\varsigma_j) - \mathbf{K}_s(\zeta_j; x,) \overline{\mathbf{1}}_{z,s}(\varsigma_j)] F_\xi(d\xi)]^2$. Consequently

$$V_n \xrightarrow{d} N(0, \sigma_0^2). \tag{A.5}$$

For R_n, it is easy to verify that $E(R_n) = 0$ and $E(R_n^2) = O(n(h!)^{-1}) = o(1)$. So $R_n = o_P(1)$ by the Chebyshev inequality. Combined with (A.4) and (A.5), we have

$$D_{n21} - B_n \xrightarrow{d} N(0, \sigma_0^2). \tag{A.6}$$

Let $b_{rs}(\varsigma_i) \equiv e_1' \{[\overline{\mathbf{S}}_r(X_i)]^{-1} \overline{\mathbf{B}}_r(\varsigma_i) - [\overline{\mathbf{S}}_s(X_i)]^{-1} \overline{\mathbf{B}}_s(\varsigma_i)\}$. Then $D_{n22} = \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} (D_{n22,rs1} - D_{n22,rs2})$, where $D_{n22,rs1} \equiv (h!)^{1/2} \sum_{i=1}^n e_1' [\overline{\mathbf{S}}_r(X_i)]^{-1} \mathbf{V}_{nr}(\varsigma_i) b_{rs}(\varsigma_i)$ and $D_{n22,rs2} \equiv (h!)^{1/2} \sum_{i=1}^n e_1' [\overline{\mathbf{S}}_s(X_i)]^{-1} \mathbf{V}_{ns}(\varsigma_i) b_{rs}(\varsigma_i)$. Write

$$\begin{aligned} D_{n22,rs1} &= n^{-1}(h!)^{1/2} \sum_{i=1}^n \sum_{j \ne i}^n e_1' [\overline{\mathbf{S}}_r(X_i)]^{-1} K_{h\lambda}(X_j, X_i) \mathbf{1}_j^r \boldsymbol{\tau}_h(X_j^c - X_i^c) \overline{\mathbf{1}}_{Z_i,r}(\varsigma_j) b_{rs}(\varsigma_i) \\ &\quad + n^{-1}(h!)^{1/2} \sum_{i=1}^n e_1' [\overline{\mathbf{S}}_r(X_i)]^{-1} K_{h\lambda}(X_i, X_i) \mathbf{1}_i^r \boldsymbol{\tau}_h(X_i^c - X_i^c) \overline{\mathbf{1}}_{Z_i,r}(\varsigma_i) b_{rs}(\varsigma_i) \\ &\equiv D_{n22,rs1a} + D_{n22,rs1b}, \text{ say.} \end{aligned}$$

Noting that $b_{rs}(\varsigma_i) = O_P(\|h\|^2 + \|\lambda\|)$, it is straightforward to show that $D_{n22,rs1b} = O_P((h!)^{-1/2} (\|h\|^2 + \|\lambda\|)) = o_P(1)$. Noting that $E(D_{n22,rs1a}) = 0$ and $E(D_{n22,rs1a}^2) = O(nh!(\|h\|^2 + \|\lambda\|)^2) = o(1)$, we have $D_{n22,rs1a} = o_P(1)$ by the Chebyshev inequality. Similarly, we can show that $D_{n22,rs1b} = o_P(1)$ and thus $D_{n22,rs1} = o_P(1)$. By the same token, $D_{n22,rs2} = o_P(1)$. It follows that

$$D_{n22} = o_P(1). \tag{A.7}$$

By Lemma A.1 and Assumption A.5, we have $D_{n23} = n(h!)^{1/2} O_P(\|h\|^4) = O_P(n\|h\|^4 (h!)^{1/2}) = o_P(1)$. This, in conjunction with (A.3), (A.6) and (A.7), implies that $D_{n2} - B_n \xrightarrow{d} N(0, \sigma_0^2)$.

Next, we show (iii). By Lemma A.1, under $H_1\left(n^{-1/2}(h!)^{-1/4}\right)$ we have

$$\begin{aligned}
D_{n3} &= \sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1} n^{-1/2}\,(h!)^{1/4}\sum_{i=1}^{n}\left[e_1'[\overline{\mathbf{S}}_r\,(X_i)]^{-1}\mathbf{V}_{nr}\,(\varsigma_i)-e_1'[\overline{\mathbf{S}}_s\,(X_i)]^{-1}\mathbf{V}_{ns}\,(\varsigma_i)\right)\\
&\quad + e_1'[\overline{\mathbf{S}}_r\,(X_i)]^{-1}\overline{\mathbf{B}}_r\,(\varsigma_i)-e_1'[\overline{\mathbf{S}}_s\,(\zeta_i)]^{-1}\overline{\mathbf{B}}_s\,(\varsigma_i)\big]\,\delta_{n,rs}(\varsigma_i)\\
&\quad + n^{1/2}\,(h!)^{1/4}\,O_P\left(\nu_n^2+\nu_n\left(\|h\|^2+\|\lambda\|\right)\right)\\
&\equiv \sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\left[D_{n31,rs}-D_{n32,rs}+D_{n33,rs}-D_{n34,rs}\right]+o_P\,(1)\,,
\end{aligned}$$

where, for example, $D_{n31,rs}\equiv n^{-1/2}\,(h!)^{1/4}\sum_{i=1}^{n}e_1'[\overline{\mathbf{S}}_r\,(X_i)]^{-1}\;\mathbf{V}_{nr}\,(\varsigma_i)\,\delta_{n,rs}(\varsigma_i)$, and $D_{n3l,rs}$, $l=2,3,4$, are analogously defined. Decompose

$$\begin{aligned}
D_{n31,rs} &= n^{-3/2}\,(h!)^{1/4}\sum_{i=1}^{n}\sum_{j\neq i}^{n}e_1'[\overline{\mathbf{S}}_r\,(X_i)]^{-1}\boldsymbol{\tau}_h\left(X_j^c-X_i^c\right)K_{h\lambda}\,(X_j,X_i)\,\overline{\mathbf{1}}_j^r\overline{\mathbf{1}}_{Z_i,r}\,(\varsigma_j)\,\delta_{n,rs}(\varsigma_i)\\
&\quad +n^{-3/2}\,(h!)^{1/4}\sum_{i=1}^{n}e_1'[\overline{\mathbf{S}}_r\,(X_i)]^{-1}\boldsymbol{\tau}_h\,(0)\,K_{h\lambda}\,(X_i,X_i)\,\overline{\mathbf{1}}_i^r\overline{\mathbf{1}}_{Z_i,r}\,(\varsigma_i)\,\delta_{n,rs}(\varsigma_i)\\
&\equiv D_{n31,rs1}+D_{n31,rs2},\ \text{say.}
\end{aligned}$$

It is easy to show that $D_{n31,rs2}=O_P(n^{-1/2}\,(h!)^{-3/4})=o_P\,(1)$ by Assumption A.5. For $D_{n31,rs1}$, noting that $E\,[D_{n31,rs1}]=0$ and

$$\begin{aligned}
&E\,[D_{n31,rs1}]^2\\
&= n^{-3}\,(h!)^{1/2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j\neq i,i'}^{n}E\left\{e_1'[\overline{\mathbf{S}}_r\,(X_i)]^{-1}\boldsymbol{\tau}_h\left(X_j^c-X_i^c\right)K_{h\lambda}\,(X_j,X_i)\,\mathbf{1}_j^r\overline{\mathbf{1}}_{Z_i,r}\,(\varsigma_j)\,\delta_{n,rs}(\varsigma_i)\right.\\
&\quad \left.\times e_1'[\overline{\mathbf{S}}_r\,(X_{i'})]^{-1}\boldsymbol{\tau}_h\left(X_j^c-X_{i'}^c\right)K_{h\lambda}\,(X_j,X_{i'})\,\overline{\mathbf{1}}_{Z_{i'},r}\,(\varsigma_j)\,\delta_{n,rs}(\varsigma_{i'})\right\}\\
&\quad +n^{-3}\,(h!)^{1/2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}E\left\{e_1'[\overline{\mathbf{S}}_r\,(X_i)]^{-1}\boldsymbol{\tau}_h\left(X_j^c-X_i^c\right)K_{h\lambda}\,(X_j,X_i)\,\mathbf{1}_j^r\overline{\mathbf{1}}_{Z_i,r}\,(\varsigma_j)\,\delta_{n,rs}(\varsigma_i)\right.\\
&\quad \left.\times e_1'[\overline{\mathbf{S}}_r\,(X_j)]^{-1}\boldsymbol{\tau}_h\left(X_i^c-X_j^c\right)K_{h\lambda}\,(X_i,X_j)\,\mathbf{1}_i^r\overline{\mathbf{1}}_{Z_j,r}\,(\varsigma_i)\,\delta_{n,rs}(\varsigma_j)\right\}\\
&= O\left((h!)^{1/2}+n^{-1}\,(h!)^{-1/2}\right)=o\,(1)\,,
\end{aligned}$$

we have $D_{n31,rs1}=o_P\,(1)$ by the Chebyshev inequality. Hence $D_{n31,rs}=o_P\,(1)$. Similarly $D_{n32,rs}=o_P\,(1)$. Noting that $\sup_\varsigma\left|\overline{\mathbf{B}}_r\,(\varsigma)\right|=O(\|h\|^2+\|\lambda\|)$, we have

$$D_{n33,rs}\leq n^{1/2}\,(h!)^{1/4}\,O\left(\|h\|^2+\|\lambda\|\right)n^{-1}\sum_{i=1}^{n}|\delta_{n,rs}(\varsigma_i)|=O_P\left(n^{1/2}\,\|h\|^2\,(h!)^{1/4}\right)=o_P\,(1)\,.$$

Similarly $D_{n34,rs} = o_P(1)$. Consequently, $D_{n3} = o_P(1)$.

We now show (iv). Noting that $a^2 - b^2 = (a-b)^2 + 2(a-b)b$, we have $\widehat{B}_n - B_n = \sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1} B_{n1,rs} + 2\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1} B_{n2,rs}$, where

$$\begin{aligned}
B_{n1,rs} &\equiv \frac{(h!)^{1/2}}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left\{\widehat{\alpha}_{ij,r} - \widehat{\alpha}_{ij,s}\right\}^2, \\
B_{n2,rs} &\equiv \frac{(h!)^{1/2}}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}[\widehat{\alpha}_{ij,r} - \widehat{\alpha}_{ij,s}][\mathbf{K}_r\left(\zeta_j; X_i\right)\overline{\mathbf{1}}_{Z_i,r}\left(\varsigma_j\right) - \mathbf{K}_s\left(\zeta_j; X_i,\right)\overline{\mathbf{1}}_{Z_i,s}\left(\varsigma_j\right)],
\end{aligned}$$

and $\widehat{\alpha}_{ij,r} = e_1'[\mathbf{S}_{nr}(X_i)]^{-1}\boldsymbol{\tau}_h\left(X_j^c - X_i^c\right)K_{h\lambda}(X_j, X_i)\mathbf{1}_j^r\widehat{\mathbf{1}}_{Z_i,r}(\varsigma_j) - \mathbf{K}_r\left(\zeta_j, X_i\right)\overline{\mathbf{1}}_{Z_i,r}(\varsigma_j)$. Noting that $[\mathbf{S}_{nr}(X_i)]^{-1} = [\overline{\mathbf{S}}_r(X_i)]^{-1} + O_P(\nu_n)$ and $\widehat{\mathbf{1}}_{z,r}(\varsigma_j) - \overline{\mathbf{1}}_{z,r}(\varsigma_j) = F(z|X_j, r) - \widehat{F}(z|X_j, r) = O_P(\nu_n + \|h\|^2 + \|\lambda\|)$ uniformly in X_j and z, we have $\widehat{\alpha}_{ij,r} = e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1}\boldsymbol{\tau}_h\left(X_j^c - X_i^c\right)K_{h\lambda}(X_j, X_i) \times\mathbf{1}_j^r\{\widehat{\mathbf{1}}_{Z_i,r}(\varsigma_j) - \overline{\mathbf{1}}_{Z_i,r}(\varsigma_j)\} + O_P(\nu_n)$. It follows that

$$\begin{aligned}
|B_{n1,rs}| &\leq \frac{(h!)^{1/2}}{(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left\|\boldsymbol{\tau}_h\left(X_j^c - X_i^c\right)K_{h\lambda}(X_j, X_i)\right\|^2 \times O_P\left(\left(\nu_n + \|h\|^2 + \|\lambda\|\right)^2\right) \\
&= O_P\left((h!)^{-1/2}\left(\nu_n^2 + \|h\|^4 + \|\lambda\|^2\right)\right) = o_P(1),
\end{aligned}$$

and similarly $|B_{n2,rs}| = O_P\left((h!)^{-1/2}\left(\nu_n + \|h\|^2 + \|\lambda\|\right)\right) = o_P(1)$ under Assumption A.5. Consequently, $\widehat{B}_n - B_n = o_P(1)$.

For (v), noticing that

$$\begin{aligned}
\widehat{\beta}_{ij,rs} &= \frac{1}{n}\sum_{l=1}^{n}\left\{\mathbf{K}_r(\zeta_i, X_l)\overline{\mathbf{1}}_{Z_l,r}(\varsigma_i) - \mathbf{K}_s(\zeta_i, X_l)\overline{\mathbf{1}}_{Z_l,s}(\varsigma_i)\right\} \\
&\quad \times\left\{\mathbf{K}_r\left(\zeta_j, X_l\right)\overline{\mathbf{1}}_{Z_l,r}(\varsigma_j) - \mathbf{K}_s\left(\zeta_j, X_l\right)\overline{\mathbf{1}}_{Z_l,s}(\varsigma_j)\right\} + o_P(1) \\
&= \int\left\{\mathbf{K}_r(\zeta_i, x)\overline{\mathbf{1}}_{z,r}(\varsigma_i) - \mathbf{K}_s(\zeta_i, x)\overline{\mathbf{1}}_{z,s}(\varsigma_i)\right\}\{\mathbf{K}_r\left(\zeta_j, x\right)\overline{\mathbf{1}}_{z,r}(\varsigma_j) \\
&\quad - \mathbf{K}_s\left(\zeta_j, x\right)\overline{\mathbf{1}}_{z,s}(\varsigma_j)\}F_\xi(d\xi) + o_P(1),
\end{aligned}$$

we have $\widehat{\sigma}_n^2 = \sigma_0^2 + o_p(1)$ by the law of large numbers for U-statistics. ■

Proof of Theorems 4.3

Using the notation defined in the proof of Theorem 4.2, we again write $n^{-1}D_n = n^{-1}(h!)^{-1/2}(D_{n1} + D_{n2} + 2D_{n3})$. Under H_1, it is easy to show that $n^{-1}(h!)^{-1/2}D_{n1} = \sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1} E[F(Z_i|X_i, r) - F(Z_i|X_i, s)]^2 + o_P(1)$, $n^{-1}(h!)^{-1/2}D_{n2} = O_P(\nu_n^2 + \|h\|^4 + \|\lambda\|^2) = o_P(1)$, and $n^{-1}(h!)^{-1/2} D_{n3} = O_P(\nu_n + \|h\|^2 + \|\lambda\|) = o_P(1)$. On the other hand, $n^{-1}(h!)^{-1/2}\widehat{B}_n = O_P(n^{-1}) = o_P(1)$ and $\widehat{\sigma}_n^2 = \sigma_0^2 + o_P(1)$. It follows that $n^{-1}(h!)^{-1/2} T_n = (n^{-1}D_n - n^{-1}(h!)^{-1/2}\widehat{B}_n)/\sqrt{\widehat{\sigma}_n^2} \xrightarrow{P} \sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1} E[F(Z_i|X_i,r) - F(Z_i|X_i,s)]^2/\sigma_0$, and the conclusion follows. ■

References

Aitchison, J., Aitken, C. G. G., 1976. Multivariate binary discrimination by the kernel method. Biometrika 63, 413-420.

Akerlof, G. 1970. The market for lemons: quality uncertainty and the market mechanism. Quarterly Journal of Economics 84, 488-500.

Boente, G., Fraiman, R., 1991. Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes. International Statistical Review 59, 355-372.

Chiappori, P.-A., Heckman, J. J., 1999. Testing for moral hazard using dynamic data. Manuscript, Dept. of Economics, Univ. of Chicago.

Chiappori, P.-A., Salanié, B. 1997. Empirical contract theory: the case of insurance data. European Economic Review 41, 943-950.

Chiappori, P.-A., Salanié, B., 2000. Testing for asymmetric information in insurance markets. Journal of Political Economy 108, 56-78.

Chiappori, P.-A., Jullien, B., Salanié, B., Salanié, F., 2006. Asymmetric information in insurance: general testable implications. Rand Journal of Economics 37, 783-798.

Cohen, A., 2005. Asymmetric information and learning: evidence from the automobile insurance market. The Review of Economics and Statistics 87, 197-207.

Cohen, A., Siegelman, P., 2010. Testing for adverse selection in insurance markets. Journal of Risk and Insurance 77, 39-84.

Crocker, K.-J., Snow, A., 2000. The theory of risk classification. In: G. Dionne, editor, Handbook of Insurance, Chapter 8, 245-276. Kluwer Academic Publishers.

Delgado, M. A., González-Manteiga, W., 2001. Significance testing in nonparametric regression based on the bootstrap. Annals of Statistics 29, 1469-1507.

Dionne, G., Doherty, N., Fombaron, N., 2000. Adverse selection in insurance markets. In: G. Dionne, editor, Handbook of Insurance, Chapter 7, 185-244. Kluwer Academic Publishers.

Dionne, G., Gourieroux C., Vanasse, C. 2001. Testing for evidence of adverse selection in the automobile insurance market: A comment. The Journal of Political Economy 109, 444-453.

Dionne, G., Michaud, P.-C., Dahchour, M., 2004. Separating moral hazard from adverse selection in automobile insurance: Longitudinal evidence from France. Cahiers de recherche 0420, CIRPEE.

Gourieroux, C., Monfort, A., Renault, E., Trognon, A., 1987. Generalised residuals. Journal of Econometrics 34, 5-32.

Hall, P., 1984. Central limit theorem for integrated square error of multivariate nonparametric density estimators. Journal of Multivariate Analysis 14, 1-16.

Hansen, B. E., 2008. Uniform convergence rates for kernel estimation with dependent data. Econometric Theory 24, 726-748.

Huang, M., White, H., 2009. A flexible test for conditional independence. Working Paper, Dept. of Economics, UCSD.

Huang, T.-M., 2009. Testing conditional independence using maximal nonlinear conditional independence. Working Paper, National Chengchi University.

Jullien, B., Salanié, B., Salanié, F., 2007. Screening risk averse agents under moral hazard. Economic Theory 30, 151-169.

Kim, H., Kim, D., Im, S., Hardin, J. W., 2009. Evidence of asymmetric information in the automobile insurance market: dichotomous versus multinomial measurement of insurance coverage. The Journal of Risk and Insurance 76, 343-366.

Li, Q., Racine, J., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton and Oxford.

Li, Q., Racine, J., 2008. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. Journal of Business and Economic Statistics 26, 423-434.

Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency rates. Journal of Time Series Analysis 17, 571-599.

Paparoditis, E., Politis, D. N., 2000. The local bootstrap for kernel estimators under general dependence conditions. Annals of the Institute of Statistical Mathematics 52, 139-159.

Puelz, R., Snow, A., 1994. Evidence on adverse selection: Equilibrium signaling and cross-subsidization in the insurance market. The Journal of Political Economy 102, 236-257.

Racine, J., Li, Q., 2004. Nonparametric estimation of regression functions with both categorical and continuous data, Journal of Econometrics 119, 99-130.

Richaudeau, D., 1999. Automobile insurance contracts and risk of accident: an empirical test using french individual data. The Geneva Papers on Risk and Insurance Theory 24, 97-114.

Robinson, P. M., 1991. Consistent nonparametric entropy-based testing. Review of Economic Studies 58, 437-453.

Rothschild, M., Stiglitz, J. E., 1976. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. The Quarterly Journal of Economics 90, 630-649.

Saito, K., 2009. Testing for asymmetric information in the automobile insurance market under rate regulation. The Journal of Risk and Insurance 73, 335-356.

Song, K., 2009. Testing conditional independence via Rosenblatt transform. Annals of Statistics 37, 4011-4045.

Su, L., White, H., 2007. A consistent characteristic-function-based test for conditional independence. Journal of Econometrics 141, 807-834.

Su, L., White, H., 2008. A nonparametric Hellinger metric test for conditional independence. Econometric Theory 24, 829-864.

Su, L., White, H., 2010. Local polynomial quantile regression for dependent heterogeneous processes: uniform Bahadur representation with applications to testing conditional independence. Working Paper, Dept. of Economics, UCSD.

Winter, R., 2000. Optimal insurance under moral hazard. In: G. Dionne, editor, Handbook of Insurance, Chapter 6, 155-184. Kluwer Academic Publishers.